Journal of Archaeological Science 30 (2003) 1695–1700

Journal of Archaeological SCIENCE

http://www.elsevier.com/locate/jas

Discussion

Comment on 'A Bayesian approach to ageing perinatal skeletal material from archaeological sites: implications for the evidence for infanticide in Roman Britain' by R.L. Gowland and A.T. Chamberlain

Simon Mays*

Ancient Monuments Laboratory, English Heritage Centre for Archaeology, Fort Cumberland, Eastney, Portsmouth PO4 9LD, UK

Received 13 August 2002; received in revised form 2 April 2003

Abstract

In 1993, I published a paper noting that the age distribution of perinatal infants for some Romano-British sites did not conform to a natural mortality pattern but rather showed a pronounced peak at a gestational age corresponding approximately to a full term infant. I interpreted this as suggestive of infanticide, given that the deed is generally carried out immediately after birth. Gowland and Chamberlain have recently published in this journal (J. Archaeol. Sci. 29 (2002) 677) a reconsideration of the problem of Roman infanticide in which they suggest that the peak I observed in the Romano-British perinatal age at death distribution may have been an artefact of the particular ageing technique I used, and they hence call into question the evidence for Roman infanticide. In this comment I argue that their work is seriously flawed and, using a re-analysis of my 1993 data, I demonstrate that the perinatal peak I observed in the Romano-British age distribution is a robust result that supports an interpretation of infanticide.

Keywords: Roman infanticide; Perinatal age estimation; Human osteology

In a paper published in 1993, I noted that the estimated age at death distribution for some Roman period perinatal infant burials differed from that of infants from a British Mediaeval site [11]. The former showed a pronounced peak at a gestational age range corresponding approximately to full term (full term is normally about 38–41 weeks gestation—Ref. [19, p. 43]), whereas the latter did not. The rather flat Mediaeval distribution was as expected for natural deaths during the perinatal period, but the Roman one was not. I interpreted the Roman distribution as suggestive of the regular practice of infanticide, given that the deed is normally carried out immediately after birth. My age estimations were based on long-bone lengths using the linear regression equations published by Scheuer et al. [16]. Scheuer and co-workers present two sets of regression equations, each derived by regressing age on bone length. One set is based on data from subjects from the Bristol Royal Hospital for Sick Children (referred to by Scheuer et al. as the BCH data), the other on data from London University Institute of Child Health (the ICH data). For my 1993 paper, I used the equations from the BCH data. These were selected because Scheuer et al. [16] provide graphs of the relationship between age and the lengths of some long bones for this sample.

Recently, in this journal, Gowland and Chamberlain [5] presented a reconsideration of the evidence for Roman infanticide. Firstly, they contend that the neonatal peak which I observed in the Roman perinatal infant age distribution may be a statistical artefact. Gowland and Chamberlain (hereafter G and C) argue this on the basis that the distribution of estimated ages in a target population (in this instance the archaeological group) may tend to mimic that of the reference sample (in this case the Scheuer et al.'s [16] BCH data) from which the regression equations used to age the target sample were derived when the ageing method is

* Tel.: +44-2392-856-779; fax: +44-2392-856-701.
E-mail address: simon.mays@english-heritage.org.uk (S. Mays).

0305-4403/03/$ - see front matter
doi:10.1016/S0305-4403(03)00053-0

derived by regressing age on age indicator [3]. This effect has been termed ‘age-structure mimicry’ [12]. The age distribution of Scheuer et al.’s [16] BCH data does indeed show a peak at approximately full term gestation, so G and C suggest the peak I reported in Roman data may be an example of the age-structure mimicry phenomenon. They then proceed to collate from the literature a different set of reference data relating bone lengths to age. They use this to re-age a Roman perinatal sample consisting of my own plus other data, and conduct a statistical manipulation using a Bayesian approach in an attempt to remove the perceived bias which, they believe, is likely to arise as a result of age-structure mimicry. The age distributions produced by these procedures still appear to show peaks at age ranges corresponding to about full term, but this effect is less pronounced.

This comment does not try to provide a general discussion of the merits of traditional regression methods versus Bayesian approaches to palaeodemography, although a properly balanced discussion of these issues is sorely needed. The aims are more modest. Firstly, I attempt to evaluate empirically Gowland and Chamberlain’s [5] claim that the neonatal peak I observed in Roman perinatal samples is a statistical artefact. Secondly, I consider critically the methodology they use to re-assess perinatal age at death in Romano-British material.

In order to investigate whether the neonatal peak I observed in my Roman data is, as Gowland and Chamberlain suggest, simply an artefact of the age distribution of Scheuer et al.’s [16] reference sample, I re-aged the material I examined for my 1993 paper using regression equations derived from another reference sample. For this I chose the radiographic study of Sherwood et al. [17]. This was a study of 136 foetuses of documented gestational age, of which 72 were morphologically normal. Sherwood and co-workers used this sample of normal infants to generate regression equations enabling estimation of age from a number of bone dimensions, including diaphysial lengths of long bones. Those for the femur, humerus, tibia, ulna and radius were used to re-age the same Roman and the Mediaeval samples previously aged [11] using Scheuer et al.’s [16] method. The femur was used when it was present, replicating the strategy used earlier with Scheuer et al.’s equations. Sherwood et al.’s [17] study was chosen because their reference sample has a very different age structure to Scheuer et al.’s [16] BCH material (Fig. 1). The age at death distributions obtained using the Sherwood et al. [17] and the Scheuer et al. [16] BCH equations on the Roman and Mediaeval infants are shown in Fig. 2.

Re-aging using Sherwood et al.’s [17] equations makes little difference to the overall shapes of the two distributions, the Romano-British data still show a

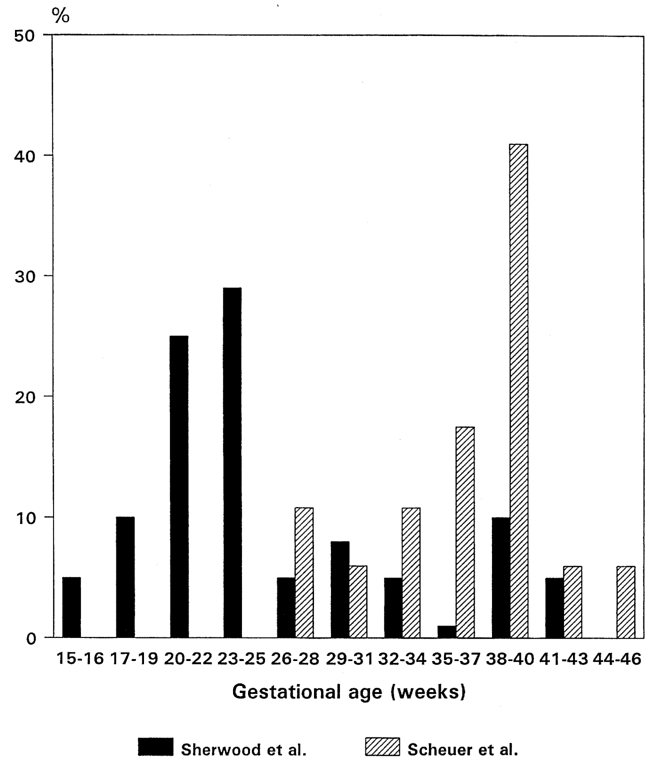

Fig. 1. Age distributions of the reference samples of Sherwood et al. [17] (N=63) and Scheuer et al. [16] (BCH data, N=17).

pronounced peak at around full term (38–41 weeks gestation), whereas the Mediaeval data are flatter. The pattern in the Roman data, whether aged using the Scheuer et al. [16] or the Sherwood et al. [17] equations, resembles closely the gestational age distribution of modern live births (Fig. 3). As I have argued previously [11], this is consistent with the regular practice of infanticide given that the deed is usually done directly after the baby is born. The Roman and Mediaeval age distributions are significantly different from one another whether aged using the Scheuer et al. [16] or the Sherwood et al. [17] equations (Kolmogorov–Smirnov D statistic: D_{obs}=0.24 if all ages are estimated using Scheuer et al.’s method [16], 0.27 if aged using Sherwood et al.’s method [17]; $D_{95\%}$=0.20).

The only difference between the ages generated from Scheuer et al.’s [16] equations and those derived using Sherwood et al. [17] is that the latter tend to be systematically older by about 1 week. This is evident as a slight ‘shuffling’ to the right of the age distributions generated by the Sherwood et al. [17] equations compared with those from the Scheuer et al. [16] method (Fig. 2). This cannot be an artefact of the different age distributions of the reference samples as that from Sherwood et al.’s [17] study is weighted toward the younger ages (Fig. 1)—if age-structure mimicry was an important effect one would expect the ages estimated

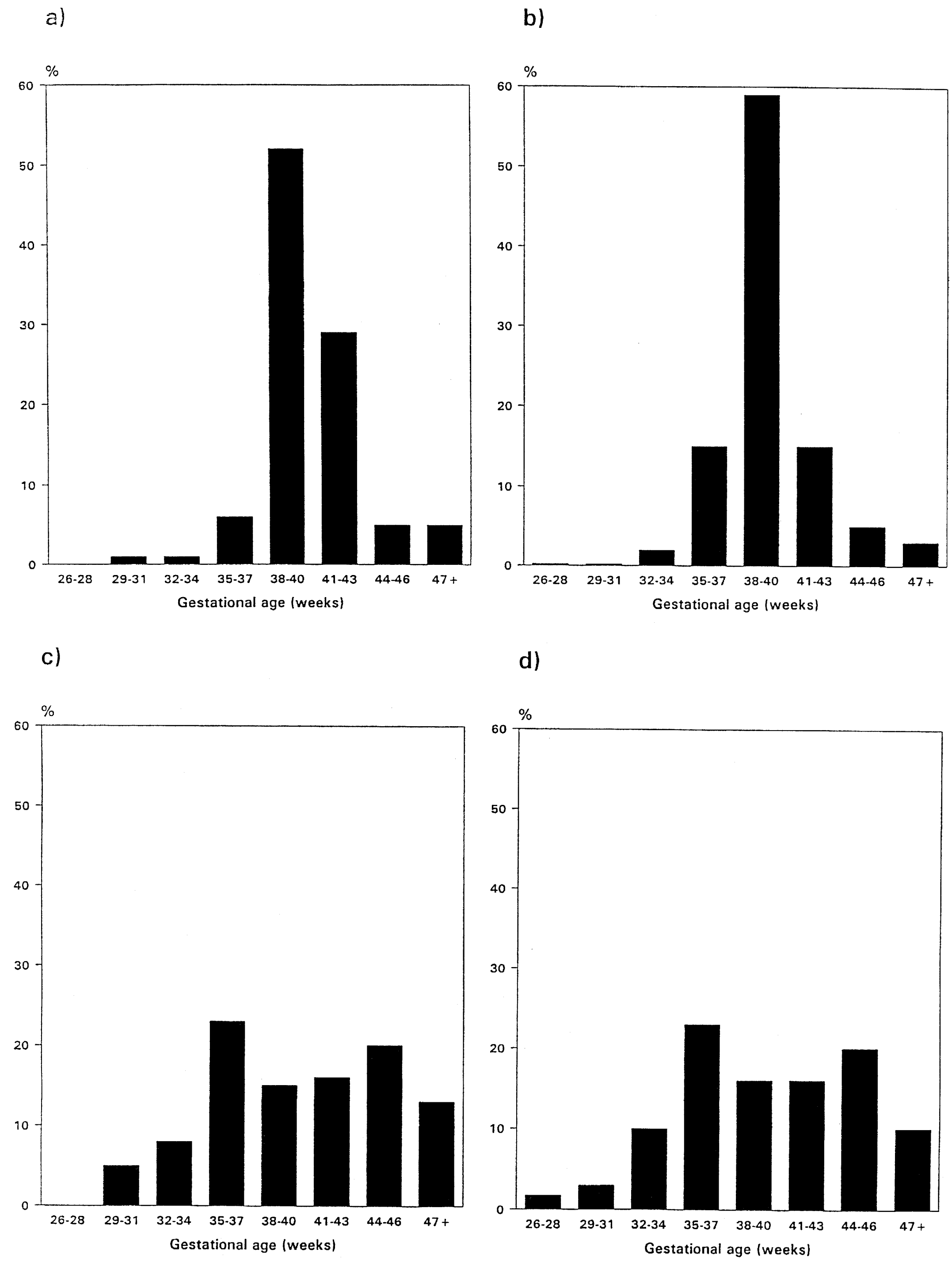

Fig. 2. Estimated perinatal age distributions for the archaeological data of Mays [11]. (a) Roman infants aged using the equations of Sherwood et al. [17]; (b) Roman infants aged using equations of Scheuer et al. [16]; (c) Mediaeval infants aged using the equations of Sherwood et al. [17]; and (d) Mediaeval infants aged using equations of Scheuer et al. [16]. For Roman data N=164, for Mediaeval data N=61. For details of sites see Mays [11].

using Sherwood et al.'s [17] equations to be biased downwards towards the reference sample mean age [7], the reverse of what is actually observed. Whether the systematic difference in ages estimated using the two techniques reflects some difference in methodology between the Scheuer et al. [16] and the Sherwood et al. [17] study or is a genuine population difference between their samples is unclear.

The above results demonstrate that the peak I observed in the Roman data is not an artefact of using the Scheuer et al.'s [16] regression equations for age estimation.

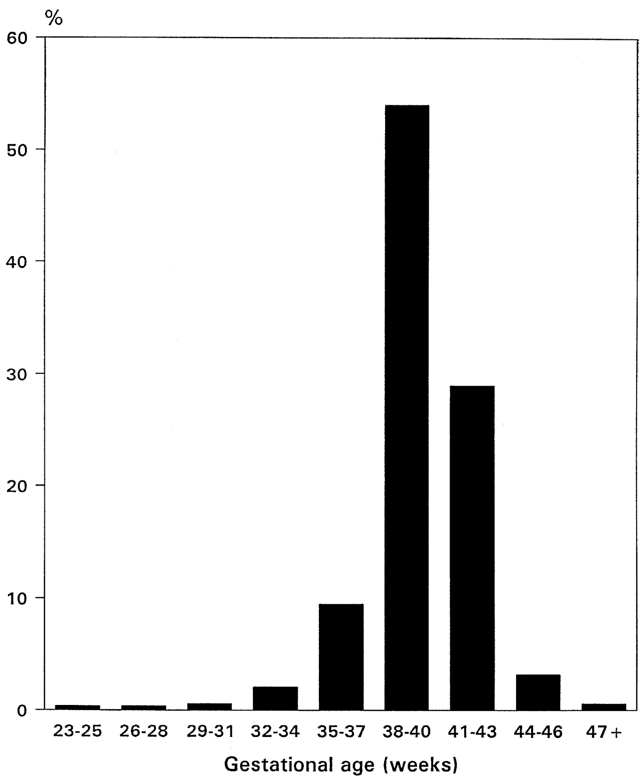

Fig. 3. Distribution of ages of total modern livebirths (N=820,532; source: Hoffman et al. [6]).

In an attempt to provide empirical support for their contention that production of a spurious perinatal peak due to age-structure mimicry of the Scheuer et al.'s [16] reference sample may be a general problem, G and C cite two further studies, in addition to my Roman data, where they claim the use of the Scheuer et al. [16] equations has produced a perinatal peak in the target sample. The other samples they cite are from Roman Ashkelon, Israel [11, Fig. 1g], and from a study of Arikara Indians [14]. The observation that the Arikara perinatal age distribution resembles the age distribution of the Scheuer BCH data is of little worth as Owsley and Jantz's [14] publication does not specify whether they used the BCH or ICH equations for age estimation (Indeed a later commentary on the Owsley and Jantz paper [7, p. 79] seems to indicate that they in fact used the ICH data.) At Ashkelon, there was other evidence, quite independent of the demographic data, which was suggestive of infanticide [18]. This indicates that infanticide rather than age-structure mimicry of Scheuer et al.'s [16] BCH reference sample may be the better explanation for the perinatal age at death profile in that instance. Curiously, in this part of their discussion, G and C fail to mention my Mediaeval age distribution ([11, Fig. 1c] reproduced as Fig. 2d in the present paper) which was, like my Roman data, based on age estimations using Scheuer et al.'s [16] BCH equations. The Mediaeval age data do not support G and C's age-structure mimicry hypothesis, as the distribution shows no strong neonatal peak and, as I have demonstrated, it differs significantly from my Roman age distribution.

This lack of empirical evidence to support the idea that age-structure mimicry of the reference sample is a general problem for perinatal age estimation from long-bone lengths using traditional regression techniques may not in fact be very surprising. Citing Aykroyd et al. [1], G and C state that the poorer the correlation between a skeletal age indicator and age, the greater the degree to which the inferred ages in the target sample will resemble the age distribution of the reference sample. They claim that long-bone length in particular has a poor correlation with age [5, p. 678]. This latter statement is untrue for foetal and perinatal infants. The most widely studied long-bone dimension in foetal and perinatal individuals of known age is femur length. Among well-founded studies, the lowest correlation coefficient I could find reported between femur length and age is 0.78 (for Scheuer et al.'s [16] ICH data). However, this figure is exceptional, correlation coefficients of around 0.95 are routinely reported in the literature, both from radiographic and ultrasound studies. For example, the radiographic studies of Scheuer et al. [16] on their BCH data, and of Sherwood et al. [17], found correlations of 0.95 and 0.96, respectively; the ultrasound investigations of Yeh et al. and Warda et al. [21,22] found coefficients of 0.95 and 0.98, respectively.

Although the main problem that I can see with G and C's work is that they fail to substantiate the assertion that the neonatal peak which was observed in the Mays' [11] Roman perinatal age data is an artefact of the Scheuer et al. [16] ageing method, there also seem to be difficulties with the reference data which they assemble in order to provide an alternative basis for ageing archaeological perinatal material.

Gowland and Chamberlain produce contingency tables, of which that for the femur is shown in their publication [5, Table 3], cross-tabulating bone length data grouped into 5 mm classes with gestational age grouped into 2-week categories. These data come from six different studies using differing methodologies. G and C acknowledge that combining data generated using different methodologies will cause increased random noise in the data (over and above variation due to real population differences) but say it will be 'largely subsumed' within the 5 mm bone length categories. However, the increased random noise resulting from combining studies with differing methodologies will inevitably cause a greater spread of bone lengths for any given age, and this will likely cause greater spread of bone lengths between different length classes for any given age category than would be the case if a single reference study, or a combination of studies with mutually consistent methodologies, was used. This alone

will, to an extent, tend to increase the spread of the estimated ages in a target sample and so will tend to 'smooth out' any neonatal peak.

One of the sources for G and C's reference data is the study of foetal osteology by Fazekas and Kosa [4]. The Fazekas and Kosa [4] data are problematic for age estimation as their individuals were for the most part of undocumented age, age being estimated from crown-heel length (see also Ref. [8]). That a reference sample used to generate ageing techniques should itself comprise individuals of accurately documented age is a fundamental requirement. The deficiency of the Fazekas and Kosa [4] material in this respect precludes the use of their study for age estimation in perinatal remains. The problem is a significant one given the large sample size (N=138) of the Fazekas and Kosa study which means that their data form a large proportion of G and C's reference group (this is especially acute for the femur where the two cited ultrasound studies had to be excluded, leaving only Scheuer et al. [16] (N=17) and Fazekas and Kosa [4] (N=138) as data sources for foetal material). The problem with the Fazekas and Kosa material for ageing studies has been recognised previously (e.g. Refs. [13,15,20]). Two of the other sources of reference data used by G and C relate to post-natal infants. One is the longitudinal study of Maresh and Deming [10] in which infant bones were measured radiographically every 6 weeks from birth until sometime between 12 and 36 weeks depending on the individual. The other is the publication of Maresh [9] which provides radiographic data on the bones of infants 2 months old. Inclusion of these data by G and C in their bone length/gestational age contingency tables is problematic since the gestational ages at birth of these individuals were not known; by treating these data as if they were gestational ages G and C will falsely increase the variability in bone lengths for a given age for their older perinatal infants. Furthermore, although Maresh and Deming [10] give their raw data, Maresh [9] does not, nor does he give data in terms of 5 mm bone length classes as required for G and C's contingency tables.

In summary, although G and C [5] present an interesting reconsideration of some of the difficulties in reconstructing and interpreting perinatal age at death profiles, problems with their reference samples undermine the validity of the results of their re-assessment of Romano-British data. In any event, there is no empirical evidence that the bias they claim to have identified in perinatal age distributions generated using the Scheuer et al. [16] equations exists to any important extent. Re-ageing my previously published perinatal samples using equations based on a reference sample with a very different age distribution from that used by Scheuer et al. [16] indicated that the peak in the Romano-British age profiles (which I interpreted as reflecting infanticide) is a robust result and cannot simply be an artefact of the Scheuer et al. [16] ageing technique. Given the strong correlation between age and long-bone length in the foetal and perinatal period this is as expected and, for the same reason, one would not anticipate that the age-structure mimicry bias would generally be a significant problem for ageing perinatal infants from bone lengths using traditional regression techniques. However, additional empirical confirmation of this would clearly be welcome. For example, it would be useful to examine the performance of various regression-based and Bayesian methodologies for recovering age distributions using samples of documented age at death. Some work has been done in this direction using adult remains (e.g. Ref. [2]), but to my knowledge this has yet to be done on perinatal material.

Acknowledgement

I am grateful to Sebastian Payne for his comments on an earlier draft of this paper.

References

[1] R.G. Aykroyd, D. Lucy, A.M. Pollard, T. Solheim, Technical note: regression analysis in adult age estimation, American Journal of Physical Anthropology 104 (1997) 259–265.

[2] R.G. Aykroyd, D. Lucy, A.M. Pollard, C.A. Roberts, Nasty, brutish, but not necessarily short: a reconsideration of the statistical methods used to calculate age at death from adult human skeletal and dental age indicators, American Antiquity 64 (1999) 55–70.

[3] J-P. Bocquet-Appel, C. Masset, Farewell to palaeodemography, Journal of Human Evolution 11 (1982) 321–333.

[4] I.G. Fazekas, F. Kosa, Forensic Foetal Osteology, Akademiai Kiado, Budapest, 1978.

[5] R.L. Gowland, A.T. Chamberlain, A Bayesian approach to ageing perinatal skeletal material from archaeological sites: implications for the evidence for infanticide in Roman Britain, Journal of Archaeological Science 29 (2002) 677–685.

[6] H.J. Hoffman, C.R. Stark, F.E. Lundin, J.D. Ashbrook, Analysis of birth weight, gestational age and fetal viability, US births 1968, Obstetrical and Gynecological Survey 29 (1974) 651–681.

[7] L.W. Konigsberg, S.R. Frankenberg, R.B. Walker, Regress what on what? Palaeodemographic age estimation as a calibration problem, in: R.R. Paine (Ed.), Integrating Archaeological Demography: Multidisciplinary Approaches to Prehistoric Population, Center for Archaeological Investigations Occasional Paper No. 24, Southern Illinois University, Carbondale, 1997, pp. 64–88.

[8] F. Kosa, Age estimation from the foetal skeleton, in: M.Y. Iscan (Ed.), Age Markers in the Human Skeleton, Charles C. Thomas, Springfield, 1989, pp. 21–54.

[9] M.M. Maresh, Linear growth of long bones of extremeties from infancy through adolescence, American Journal of Diseases of Children 89 (1955) 725–742.

[10] M.M. Maresh, J. Deming, The growth of the long bones in 80 infants. Roentgenograms versus anthropometry, Child Development 10 (1939) 91–106.

[11] S. Mays, Infanticide in Roman Britain, Antiquity 67 (1993) 883–888.

[12] R.P. Mensforth, Palaeodemography of the Carlston Annis (Bt5) late Archaic skeletal population, American Journal of Physical Anthropology 82 (1990) 81–99.

[13] T.I. Molleson, Social implications of mortality patterns of juveniles from Poundbury Camp, RomanoBritish cemetery, Anthropologischer Anzeiger 47 (1989) 27–38.

[14] D.W. Owsley, R.L. Jantz, Long bone lengths and gestational age distributions of post-contact period Arikara Indian perinatal infant skeletons, American Journal of Physical Anthropology 68 (1985) 321–328.

[15] L. Scheuer, S. Black, Development and ageing of the juvenile skeleton, in: M. Cox, S. Mays (Eds.), Human Osteology in Archaeology and Forensic Science, Greenwich Medical Media, London, 2000, pp. 9–21.

[16] J.L. Scheuer, J.H. Musgrave, S.P. Evans, Estimation of late foetal and perinatal age from limb bone length by linear and logarithmic regression, Annals of Human Biology 7 (1980) 257–265.

[17] R.J. Sherwood, R.S. Meindl, H.B. Robinson, R.L. May, Fetal age: methods of estimation and effects of pathology, American Journal of Physical Anthropology 113 (2000) 305–315.

[18] P. Smith, G. Kahila, Identification of infanticide in archaeological sites: a case study from the late Roman-early Byzantine periods at Ashkelon, Journal of Archaeological Science 19 (1992) 667–675.

[19] J.M. Tanner, Foetus Into Man, second ed., Castlemead, Ware, 1989.

[20] M.W. Tocheri, J.E. Molto, Ageing foetal and juvenile skeletons from Roman period Egypt using basiocciput osteometrics, International Journal of Osteoarchaeology 12 (2002) 356–363.

[21] A.H. Warda, R.L. Deter, I.K. Rossavik, R.J. Carpenter, F.P. Hadlock, Fetal femur length: a critical reevaluation of the relationship to menstrual age, Obstetrics and Gynaecology 66 (1985) 69–75.

[22] M.-N. Yeh, L. Bracero, K.B. Reilly, L. Murtha, M. Abroulafia, B.A. Barron, Ultrasonic measurement of the femur length as an index of fetal gestational age, American Journal of Obstetrics and Gynecology 144 (1982) 519–522.